

Chapter 8

Linear Mixed-Effects and Latent Curve Models for Longitudinal Life Course Analyses

Paolo Ghisletta, Olivier Renaud, Nadège Jacot, and Delphine Courvoisier

Introduction

The core of life course or lifespan research consists in studying how individual trajectories are shaped and unfold over time, from conception to death (Baltes 1987). Two concepts are fundamental in this endeavor: stability and change. Indeed, while certain individual characteristics remain constant across one's lifespan (e.g., sex, ethnicity), others undergo profound change, to the point that they might mutate into other characteristics (e.g., health, cognitive capacities). Such changes need not be independent of each other. They may simply co-occur, in the sense that while they happen simultaneously, it is hard to infer causality mechanisms between them. They may also be intrinsically related, where one change process is a necessary antecedent of the other, which becomes the consequence.

The empirical study of lifespan development necessitates hence statistical models capable, at the very least, of (a) estimating constancy and change in information from a collection of observed measurements and (b) assessing degrees of

P. Ghisletta (✉) • N. Jacot

Faculty of Psychology and Educational Sciences, University of Geneva, Geneva, Switzerland

Distance Learning University Switzerland, Brig, Switzerland

e-mail: paolo.ghisletta@unige.ch; nadege.jacot@unige.ch

O. Renaud

Faculty of Psychology and Educational Sciences, University of Geneva, Geneva, Switzerland

e-mail: olivier.renaud@unige.ch

D. Courvoisier

Division of Clinical Epidemiology, University Hospitals of Geneva, Geneva, Switzerland

Swiss National Center of Competence for Research LIVES – Overcoming Vulnerability: Life Course Perspectives, University of Geneva, Geneva, Switzerland

e-mail: delphine.courvoisier@unige.ch

© The Author(s) 2015

C. Burton-Jeangros et al. (eds.), *A Life Course Perspective on Health Trajectories and Transitions*, Life Course Research and Social Policies 4,
DOI 10.1007/978-3-319-20484-0_8

155

interrelationships among constant and changing characteristics. Moreover, in a contextual paradigm typical of life course research, it is also highly desirable to ascertain which characteristics are intrinsic to the individual and which are influenced by, or stem from, external factors. Thus, the statistical models adopted in life course research face a number of challenging questions, which in turn can only be met with an appropriate data collection methodology.

The Rationale and Objectives of Longitudinal Research

There is common agreement that to study any changing phenomenon it is best to utilize longitudinal data, rather than cross-sectional data with only one time of measurement per individual. Baltes and Nesselroade (1979), in their seminal chapter, have defined the rationale of longitudinal research as “the study of phenomena in their time-related constancy and change” (p. 2). This conceptualization encompasses various types of methodologies applied across many disciplines, such as panel or wave designs, repeated measurement experiments, single-subject designs, and time series. The common feature of these designs is the presence of repeated observations on the same units of observation (usually individuals). For instance, the same individuals are assessed multiple times, necessarily at different time points, and the researcher’s interest lies in assessing what is constant, what changes across the multiple measurements, and what may explain the changes.

Following this rationale, Baltes and Nesselroade (1979) have outlined five objectives of longitudinal research. We repeat them here because these objectives motivate the use of longitudinal research and thereby should be directly addressed by statistical models applied to longitudinal data. To clarify, we provide a working example. The five objectives are:

- Direct identification of *intraindividual change*. This objective aims at estimating change over (at least two) repeated assessments at the individual level, thereby identifying individual trajectories of change. The estimation of intraindividual change is at the very heart of longitudinal research and can only be approximated (usually unsatisfactorily) by analyzing single measurements of different individuals assessed at different time points. Cross-sectional data, thus, do not allow the exploration of this objective. A researcher interested, for instance, in assessing social inequalities in health trajectories across the lifespan (e.g., Cullati et al. 2014) will start out by assessing whether the health status of individuals changes across time. Individuals for whom intraindividual change is observed will not display flat trajectories.
- Direct identification of *interindividual differences* (or similarities) *in intraindividual change*. In life course and lifespan research it is unrealistic to assume that entities are perfect replicates of each other. Hence, once intraindividual change is identified, this objective focuses on assessing whether individuals are similar

to or differ from each other with respect to their specific intraindividual change patterns. Indeed, while all individuals may change in time, some may display increase in the intensity of the observed behavior in time, while others may show decrease. To continue the previous example, it may well be that for some individuals their health trajectories are rather stable in time, while for others their trajectories decline, evincing deterioration in general health, or rise, signifying health improvement. Lack of interindividual differences in intraindividual change would be reflected by parallel health trajectories, meaning that all individuals' health changes equally.

- Analysis of *interrelationships in change*. It is seldom that life course researchers focus on single variables, because rarely do important characteristics of our lifespan change in a completely independent and autonomous fashion. Most often the interest lies in understanding how multiple variables are associated, either occurring in parallel, or influencing each other. For instance, change in one observed behavior (e.g., declining health) may occur in parallel with other characteristics (e.g., aging, socioeconomic status, increase in body mass index, loss of income), or be more likely to come about if another behavior is previously triggered (e.g., loss of employment). Such interrelationships, again, cannot be approximated by cross-sectional methodology, unless the characteristics are inert and do not undergo any change. That is, static correlations between cross-sectional assessments of behaviors are usually bad approximations of the dynamics at play between these behaviors.
- Analyses of *causes (determinants) of intraindividual change*. If intraindividual change can be identified (cf., first objective), it becomes highly pertinent to determine possible causes (or determinants) of such change. The passing of time is a convenient metric along which to describe and analyze change, but it seldom constitutes a theoretically sound explanation of any change process. While we know that in old age health generally declines, it is recognized that various causes are at play at very different levels (e.g., cellular, biological, epigenetic, psychological, social, all of which may correlate with aging).
- Analyses of *causes (determinants) of interindividual differences in intraindividual change*. If interindividual differences in intraindividual change (second objective) are detected, we may wonder what causes such differences. It may well be that the causes of intraindividual change (fourth objective) are different across individuals (e.g., while an individual's health improvement may primarily be due to change in marital status, another individual's health improvement might be triggered by change in occupation status). This fifth objective seems promising with regard to addressing social inequalities in health trajectories across the lifespan.

The statistical models we discuss in this chapter are widely used in many disciplines because they define parameters that address directly the five objectives of longitudinal research. Moreover, recent extensions of these models allow broadening the scopes of longitudinal research by exploring additional questions, such as the presence of interindividual differences due to (known or unknown)

group membership. These models have also been adapted to handle realities of longitudinal designs which render the whole research enterprise more arduous, such as dealing with unavoidable missing values.

A Linear Model of Change

Let Y_{ij} represent the measurement of Y at time i for individual j . Let's assume there are $j = 1, 2, \dots, n$ individuals composing a sample of size n and let's assume this sample has been assessed at times $i = 1, 2, \dots, T$, hence T times. The usual statistical assumption of such a sample is that the n individuals are independent of each other, meaning that the scores Y_{ij} of an individual j are not related to the scores $Y_{ij'}$ of another individual j' , and this at any time point i . Nevertheless, it is natural to expect that the score Y_{ij} of individual j at time i is correlated with the score $Y_{i'j}$ of that same individual at another time point i' , given that both depend on that individual's characteristics. The first objective of longitudinal research motivates understanding how the repeated measurements Y_{ij} are related in time. In other words, a mathematical model is wished for, to relate a meaningful variable that changes in time (hence assessed at multiple time points i) to the outcome of interest Y , and this for every individual j .

Historically, longitudinal data were first analyzed one individual at a time, separately, in what was named individual growth models (the term growth most probably is due to the fact that the first applications of this kind studied the physical growth of children undergoing puberty; see McArdle 2001). In other words, for n individuals, n models were computed separately. In this context, most frequently the repeated assessments were related to the aging of the individual, where a_{ij} is the age of individual j at time i . Note that other definitions of time can be specified (e.g., occasions of measurement, assessment trials in a laboratory experiment, hours of the day when studying a circadian function). A simple model assuming that Y is linearly related to age (a_{ij}) is the simple linear regression model as specified in Eq. 8.1. Because this model has to be estimated separately for each individual j , we drop the subscript j .

$$Y_i = \pi_0 + \pi_1 a_i + E_i \quad (8.1)$$

In the end, the analyst will obtain, for each individual j , a value for: (a) the intercept π_0 , which defines the predicted value of Y when age = 0; (b) the slope π_1 , which defines the predicted change (either positive for growth or negative for decline) of Y for each unit change (e.g., one year, one month) in age; (c) the standard error of the estimate, which is closely related to the variability of the errors E_i at each time point and which defines the quality of the overall age prediction of Y . Note that to interpret the intercept, it is customary to center age around a meaningful value, such as the average of the sample (by subtracting from each individual's age the sample average age). The intercept is then the predicted Y value for an individual

of average age. Alternative age scalings have been proposed and might allow for a more meaningful interpretation of the intercept in particular research situations (e.g., Mehta and West 2000; Wainer 2000). The assumptions of the model are that the errors E_i are normally distributed (i.e., they follow a normal, Gaussian curve) and do not depend on values of age (i.e., homoscedasticity).

If the model is estimated for each individual, n estimates of these parameters are obtained. Of course, other mathematical relations between age and Y can be tested, such as polynomial or exponential functions (for instance, to model human growth, various exponential functions have been proposed). What must be kept in mind is that so far the analysis is individual-specific. At the first analytical step, the growth model is estimated for each individual. At the second analytical step, the individual estimates are subsequently summarized. Any conclusion about the overall sample would have to be inferred by summarizing the n estimates, for instance by calculating the average and the variance of the intercepts and of the slopes across all individuals. Note that the two steps are computed independently of each other.

The Linear Mixed-Effects Model

In 1982 Laird and Ware proposed a model that allowed estimating simultaneously intercept and slope information at both the individual and the sample level. The model is an expansion of the individual growth model and is presented in Eq. 8.2.

$$\begin{aligned} Y_{ij} &= \pi_{0j} + \pi_{1j}a_{ij} + E_{ij} \\ \pi_{0j} &= \beta_0 + U_{0j} \\ \pi_{1j} &= \beta_1 + U_{1j} \end{aligned} \tag{8.2}$$

Given that the repeated assessments of Y of all individuals are analyzed simultaneously, Eq. 8.2 necessitates the addition of the subscript j , which identifies the individual. Moreover, this approach supposes one set of growth parameters, that is, one intercept and one slope, per individual, which again justifies the subscript j on both parameters π_0 and π_1 . Technically, it is not correct to say that an intercept and a slope value are estimated for each individual. That is, the model does not explicitly estimate a π_{0j} and a π_{1j} value for each individual j . The model presupposes, however, that each individual may have an intercept and a slope value that deviate from the central (population average) values, which are indicated by β_0 and β_1 and which are explicitly estimated. What are also estimated are the inter-individual variances, due to the individual deviations U_{0j} and U_{1j} around the central values, and possibly the covariance between U_{0j} and U_{1j} . The variances of the U_{0j} and U_{1j} are defined by the parameters σ^2_I and σ^2_S , respectively, and their covariance by the parameter σ_{IS} . Lastly, the errors of prediction E_{ij} are not individually estimated. These are often assumed to have a constant variance in time, estimated by the parameter σ^2_E , and to be uncorrelated in time. In sum, then, in its

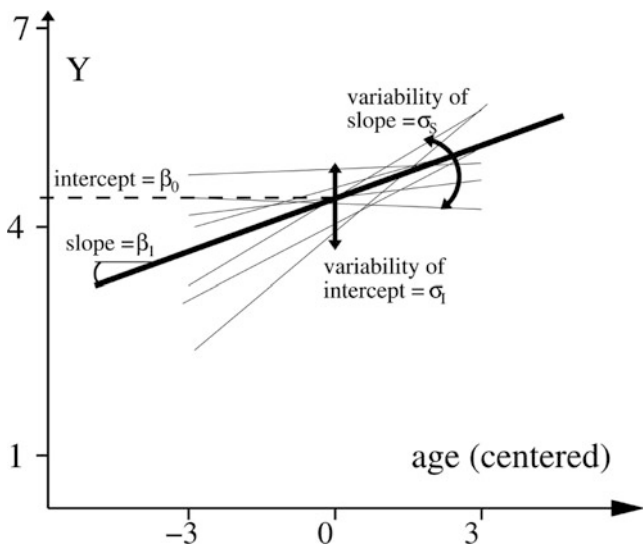


Fig. 8.1 Schematic representation of the parameters estimated in the LMEM (Eq. 8.2). The thin lines represent the best-fitting trajectory of each individual, while the thick line represents the best-fitting trajectory based on the central values (representative of the overall sample). In this example the value of the slope, β_1 , is positive. Note that the estimated slope variance, σ^2_S , and intercept-slope covariance, σ_{IS} , depend on where age has been centered

most frequent specification, this model estimates six parameters: two central values, for the intercept, β_0 , and the slope, β_1 , two variances and a covariance of growth parameters, σ^2_I , σ^2_S , and σ_{IS} , and an error variance σ^2_E . Figure 8.1 illustrates schematically the parameters of the *linear mixed-effects model* (LMEM).

This model distinguishes two kinds of parameters: the central values, which are common to all individuals (β_0 and β_1) and are called *fixed effects*, and the individual deviations from these central values (U_{0j} and U_{1j}), called the *random effects*. Again, the random effects and the individual errors (E_{ij}) are not directly estimated, but their variances and covariance (σ^2_I , σ^2_S , σ_{IS} , and σ^2_E) are. Fixed effects thus apply to all individuals and are not subject to individual variations. Random effects, on the other hand, vary across individuals and do so by typically following the standard distribution of random variables, that is, a normal (Gaussian) distribution. The model hence assumes that all random effects (U_{0j} , U_{1j}) and errors (E_{ij}) are normally distributed (symbolized by $U_{0j} \sim \mathcal{N}(0, \sigma^2_I)$, $U_{1j} \sim \mathcal{N}(0, \sigma^2_S)$, and $E_{ij} \sim \mathcal{N}(0, \sigma^2_E)$), that the random effects may covary ($\text{Cov}(U_{0j}, U_{1j}) = \sigma_{IS}$), that the errors do not covary with the random effects ($\text{Cov}(E_{ij}, U_{0j}) = \text{Cov}(E_{ij}, U_{1j}) = 0$), and that the errors do not covary in time ($\text{Cov}(E_{ij}, E_{i'j}) = 0$).

The coexistence of fixed and random effects within the same model gives it the name of the linear mixed-effects model. The term linear denotes that the function linking Y to the predictor a_{ij} is linear in its parameters, meaning that the parameters associated to the prediction of Y (on the right side of the equal sign in Eq. 8.2)

are at most multiplied by a predictor and then added (i.e., the prediction is a linear combination of the parameters and the predictors), rather than, for instance, being exponentiated. This model has been developed in different disciplines, and is also known as random-effects model (Laird and Ware 1982), hierarchical linear model (Bryk and Raudenbush 1987), and multilevel model (Goldstein 1989).

The LMEM of Eq. 8.2 has notable advantages over the individually estimated growth models of Eq. 8.1. First, the estimation is simultaneous, that is, instead of involving two analytical steps it can be computed in a single step. All n individuals' data are analyzed together, in a single analysis. This makes for a considerable gain in time. Second, the statistical tests obtained with the LMEM are superior to those of individually estimated growth models. More precisely, the Type I error rate of statistical tests is closer to its nominal value within the LMEM than in individual growth models, which are usually too liberal (Snijders and Bosker 2012). Third, the LMEM allows for a statistical test of sample heterogeneity, which is not possible with the individual growth approach. Hence, statistical tests for the significance of variances and covariance of the random effects (σ^2_I , σ^2_S , and σ_{IS}) are possible within the LMEM. This feature is extremely important, as these parameters represent heterogeneity in growth parameters, a concept often of chief interest from a theoretical perspective. Fourth, the LMEM can be extended to define parameters that operationalize the five objectives of longitudinal research: (a) Intraindividual change is directly identified with the first line of Eq. 8.2, and β_0 and β_I define the average intraindividual change function; (b) the intercept variance σ^2_I and the slope variance σ^2_S identify directly interindividual differences in intraindividual change; (c) in multivariate specifications of the LMEM it is possible to covary intercept and slope of one change process with those of another (MacCallum et al. 1997; more detail in the section “Three Notable Extensions”, below); (d) it is straightforward to add a predictor to the first line of Eq. 8.2 to test a determinant of intraindividual change; and (e) if intercept variance and slope variance are significant, it is straightforward to extend the LMEM to include a predictor of interindividual differences in intraindividual change (more detail in the section “Inclusion of Covariates”, below). In other words, the LMEM appears to provide explicit statistical tests directly associated to the rationale and objectives of longitudinal research enunciated by Baltes and Nesselrode (1979). Consequently, this statistical model addresses key theoretical questions of life course research.

The Latent Curve Model

In 1984, Meredith and Tisak 1984 presented the precepts of how the LMEM defined above can be specified as a structural equation model. The work was later formalized by the same authors (Meredith and Tisak 1990), and applied by McArdle (1986). Structural equation modeling can largely be defined as a set of statistical techniques aimed at testing hypothesized relationships among chosen variables. In this context, the Latent Curve Model (LCM) formalizes how a series of repeated measurements

of variable Y for individual j , represented by the vector \mathbf{Y}_j , is related to the passing of time (or aging of the individual) as specified in Eq. 8.3 (note that it is customary to write the names of vectors and matrices in boldface).

$$\mathbf{Y}_j = \mathbf{\Lambda} \boldsymbol{\eta}_j + \mathbf{E}_j \quad (8.3)$$

For each individual j , the T repeated measurements of Y are piled up in the vector \mathbf{Y}_j (of size $T \times 1$, corresponding to Y_{1j} to Y_{Tj} in Eq. 8.2). $\boldsymbol{\eta}_j$ represents the vector (of size 2×1) of the two growth factors (or latent variables), intercept and slope (equivalent to π_{0j} and π_{1j} in Eq. 8.2). $\mathbf{\Lambda}$ is the matrix (of size $T \times 2$) of the factor loadings associating the repeated observations to the growth factors. Note that $\mathbf{\Lambda}$ is not indexed by j , indicating that its values are the same for all individuals. Finally, \mathbf{E}_j is the vector (of size $T \times 1$, corresponding to E_{1j} up to E_{Tj} in Eq. 8.2) of time-specific errors. The loadings associating the intercept to the measurements (i.e., the first column of $\mathbf{\Lambda}$) are conveniently fixed at 1 (the multiplier of π_{0j} in Eq. 8.2). To specify linear change, the loadings of the slope factor (i.e., the second column of $\mathbf{\Lambda}$) increase linearly with time, or age. Hence, in this basic specification, the elements of the loading matrix $\mathbf{\Lambda}$ are not estimated. For instance, if the sample is observed at five ages, the loadings of the intercept would be $[1 \ 1 \ 1 \ 1 \ 1]'$ while for the slope they might be $[-2 \ -1 \ 0 \ 1 \ 2]'$, (the values of a_{ij} in Eq. 8.2; the prime symbol ' stands for transposed). This is equivalent to centering the variable age on the third value in the LMEM. To clarify, we expand the general LCM notation in an illustration below (cf. Eq. 8.7).

The crucial parameter estimates of the LCM are associated to the growth factors $\boldsymbol{\eta}$. The central values of the intercept and of the slope factor correspond to the fixed effects β_0 and β_1 of the LMEM; the variances of the two factors correspond to σ^2_I and σ^2_S and their covariance corresponds to σ_{IS} . Finally, if the error variance is constrained to be constant in time, its estimate corresponds to σ^2_E . In the end, it can be shown that, while they have originated in rather widely different fields of statistics, the LMEM and the LCM as specified here are completely equivalent (e.g., Bauer 2003; Chou et al. 1998; McArdle and Hamagami 1996).

Inclusion of Covariates

The growth model implemented via the LMEM or the LCM allows for testing sample heterogeneity in the growth factors. For instance, are individuals different from each other with respect to their intercept score? This question is operationalized by testing the null hypothesis that the intercept variance is zero ($H_0: \sigma^2_I = 0$). Likewise, we can ask whether there are individual differences in the change process, or whether the entities are different from each other with respect to their slope score. Testing the null hypothesis of zero slope variance ($H_0: \sigma^2_S = 0$) addresses this question. If there appear to be interindividual differences in intercept and in slope (i.e., if the two variances are different from zero), a natural question is whether the

intercept and the slope scores are related. This question is addressed by the null hypothesis of zero covariance ($H_0: \sigma_{IS} = 0$).

If sample heterogeneity in intercept and/or slope appears significant, we can wonder whether a given characteristic of the entity may influence the growth factors. This question can easily be addressed within the LMEM and LCM by expanding the models to include predictors. There are two kinds of predictors. Time-varying predictors are those that vary across time, such as a medical or physical measurement taken at each assessment time, and hence require not only a subscript j to determine the individual, but also a subscript i to specify their time of assessment. These predictors are also called time-dependent or level 1 and may help explain intraindividual change. Time-invariant predictors, on the other hand, are individual specific and do not change in time, such as sex. These are simply denoted with a subscript j and are also called level 2, and may help explain interindividual differences in intraindividual change.

The LMEM can be expanded to include a time-varying predictor x_{ij} and a time-invariant predictor z_j as shown in Eq. 8.4.

$$\begin{aligned} Y_{ij} &= \pi_{0j} + \pi_{1j}a_{ij} + \pi_{2j}x_{ij} + E_{ij} \\ \pi_{0j} &= \beta_{00} + \beta_{01}z_j + U_{0j} \\ \pi_{1j} &= \beta_{10} + \beta_{11}z_j + U_{1j} \end{aligned} \quad (8.4)$$

In the LCM the inclusion of a time-varying and a time-invariant predictor is shown in Eq. 8.5 (in which we use the notation that is standard in the structural equation modeling literature; e.g., Bollen and Curran 2006).

$$\begin{aligned} \mathbf{Y}_j &= \Lambda \boldsymbol{\eta}_j + B \mathbf{x}_j + \mathbf{E}_j \\ \boldsymbol{\eta}_j &= \boldsymbol{\alpha} + \boldsymbol{\Gamma} \boldsymbol{\xi}_j + \boldsymbol{\zeta}_j \end{aligned} \quad (8.5)$$

Equation 8.5 specifies that the outcome \mathbf{Y}_j is not only influenced by the growth factors $\boldsymbol{\eta}_j$, but also by an observed time-varying covariate \mathbf{x}_j , through a regression weight specified in the scalar B (equivalent to π_2 in Eq. 8.4). The vector \mathbf{x}_j has T rows and one column, while the scalar B is a single number. The growth factors $\boldsymbol{\eta}_j$ may be influenced by a time-invariant covariate $\boldsymbol{\xi}_j$ (z_j in Eq. 8.4) via the regression intercepts $\boldsymbol{\alpha}$ (β_{00} and β_{10} in Eq. 8.4) and the regression weights $\boldsymbol{\Gamma}$ (β_{01} and β_{11} in Eq. 8.4). Given that the prediction of $\boldsymbol{\eta}_j$ by $\boldsymbol{\xi}_j$ will likely not be perfect in practice, Eq. 8.5 includes the regression errors $\boldsymbol{\zeta}_j$ (U_{0j} and U_{1j} in Eq. 8.4). Again, it can be shown that the models in Eqs. 8.4 and 8.5 are equivalent.

Note that it is not necessary that a time-invariant predictor influence both the intercept and the slope. Just like in ordinary linear regression, multiple predictors, including interactions, can be added to the model. Also, the same substantive variable could be considered as time-varying (for instance, daily values of a characteristic) or as time-invariant (for example, average value of the same characteristic over multiple days). In such cases it is important to note that the meaning of the variable can change dramatically and special care must be taken when

interpreting the results (Goldstein 2011; Raudenbush and Bryk 2002; Snijders and Bosker 2012). Also, the LCM allows for more flexibility when testing covariates' effects. For instance, the effect of a time-varying predictor is typically specified to be the same at all time points in the LMEM. While relaxing this equality constraint requires some data manipulation in LMEM, this can easily be achieved in the LCM.

We expect the variability of the random effects and of the errors (i.e., σ^2_I , σ^2_S , and σ^2_E) to diminish as we add important predictors to the model. In other words, the unexplained variance of the models should decrease as we add important predictors (this effect can be tested statistically). Effect size measures, similar to the coefficient of determination (R^2) in ordinary linear regression, have been proposed to quantify these effects (e.g., Snijders and Bosker 1994).

LMEM or LCM?

Given the equivalency of the two models, when should an analyst use the LMEM and when the LCM? To a certain extent, the answer is a matter of preference. Given that the two approaches stem from different traditions, historically the statistical programs estimating these models were quite specific. Either they would estimate the LMEM (e.g., MIXED in SPSS, PROC MIXED in SAS, nlme and lme4 in R, mixed in STATA, HLM, MLwiN) or the LCM (e.g., LiSRel, AMOS, Mx, OpenMx in R, sem in STATA, EQS, Mplus). Recently, some software allow for both estimations (e.g., R, LiSRel, Mplus), although the optimization algorithms implemented in such software may not be the most efficient.

There are, however, empirical conditions, in which the implementation of one model over the other may be indicated. Ghisletta and Lindenberger (2004) described the relative advantages and disadvantages of the LMEM and the LCM approaches and suggested some practical guidelines concerning the choice between the two. In brief, the LMEM may be preferable over the LCM when (a) the data set is not balanced (i.e., not every entity has been observed at all times and at the same time points; hence, each entity j has been observed T_j times, where T_j may vary across entities) and there are many incomplete data; (b) the theoretical functional form of change is a common mathematical function (e.g., linear, polynomial, exponential); (c) the structure of the error covariance matrix is common (e.g., independent-diagonal, auto-regressive, unstructured); (d) the theoretical relation between the growth factors (i.e., intercept and slope) can only be a covariance (rather than, for instance, a regression weight); and (e) the role of a time-invariant variable, characteristic of the individual (e.g., sex, ethnicity, age at study inception), is that of predicting the growth factors, rather than for instance correlating with them.

The LCM approach may be more indicated than the LMEM approach when (a) there is a limited amount of unbalanced and incomplete data; (b) the change function is more complex than common mathematical functions (but remains linear in its

parameters) or, when unknown, may be estimated directly from the data; (c) the structure of the errors is of theoretical interest and requires flexible specifications; (d) the growth factors are not limited to simply covary but could be regressed on each other; and (e) the associations between external covariates and the growth factors are not limited to unilateral predictions.

These suggestions are meant as starting guidelines and not definite rules. Either approach has undergone notable extensions, which greatly amplifies their degree of flexibility in modeling change processes. For instance, the LMEM has been extended to allow relating the outcome (Y_{ij}) to the predictors (on the right side of Eqs. 8.2 or 8.4) by a so-called link function from the exponential family. This model is called the Generalized Linear Mixed Model (GLMM; McCulloch and Searle 2001). Even more flexibility about the relationships between the outcome and the predictors has been introduced in the context of the so-called Nonlinear Mixed-Effects Models (Davidian and Giltinan 1995). The LCM has also undergone notable extensions. In particular, when applied to multiple variables, the model has been extended to include dynamic associations between the growth factors, allowing to establish which variable, if any, is driving the change of the other variable, hence determining the leader and the lagger within a multivariate dynamic system (McArdle and Hamagami 2001).

An Illustration from the Swiss Household Panel

To illustrate the LMEM and the LCM, we use data from the second sample of the Swiss Household Panel (SHP), which is a study based at the Swiss Centre of Expertise in the Social Sciences FORS. The project is financed by the Swiss National Science Foundation. The SHP is a nationally representative survey of households investigating trends in social dynamics among the Swiss population. Data are collected annually by Computer Assisted Telephone Interviewing. For this illustration, we focus on eight waves collected on 3,665 individuals, running from 2004 to 2011, and examine the evolution of self-reported health (SRH) over time. SRH is measured by the first question of the Short-Form-36 (“In general, do you think your health is...”, rated from 1 (excellent) to 5 (poor)). Albeit simple, this measure is interesting because it predicts several important health outcomes, such as sick leave, and mortality (Berkman and Syme 1979; Halford et al. 2012).

We reverse coded SRH, such that 1 represents poor health and 5 indicates excellent health. Independent variables are sex (sex_j , coded 0 for men and 1 for women), a time-invariant predictor, and satisfaction with personal relationships ($satrel_{ij}$, rated from 0, not at all satisfied, to 10, completely satisfied), a time-varying predictor. In addition, to examine intraindividual change in SRH and allow for the influence of sex on this change, we included time ($time_{ij}$) as a time-varying predictor, yielding the following model:

LMEM Notation

$$\begin{aligned}
 SRH_{ij} &= \pi_{0j} + \pi_{1j}time_{ij} + \pi_{2}satrel_{ij} + E_{ij} \\
 \pi_{0j} &= \beta_{00} + \beta_{01}sex_j + U_{0j} \\
 \pi_{1j} &= \beta_{10} + \beta_{11}sex_j + U_{1j}
 \end{aligned} \tag{8.6}$$

LCM Notation

$$\begin{pmatrix} SRH_{1j} \\ SRH_{2j} \\ SRH_{3j} \\ SRH_{4j} \\ SRH_{5j} \\ SRH_{6j} \\ SRH_{7j} \\ SRH_{8j} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \\ 1 & 6 \\ 1 & 7 \end{pmatrix} \begin{pmatrix} \eta_{1j} \\ \eta_{2j} \end{pmatrix} + B \begin{pmatrix} satrel_{1j} \\ satrel_{2j} \\ satrel_{3j} \\ satrel_{4j} \\ satrel_{5j} \\ satrel_{6j} \\ satrel_{7j} \\ satrel_{8j} \end{pmatrix} + \begin{pmatrix} E_{1j} \\ E_{2j} \\ E_{3j} \\ E_{4j} \\ E_{5j} \\ E_{6j} \\ E_{7j} \\ E_{8j} \end{pmatrix} \tag{8.7}$$

$$\begin{pmatrix} \eta_{1j} \\ \eta_{2j} \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} + \begin{pmatrix} \Gamma_1 \\ \Gamma_2 \end{pmatrix} sex_j + \begin{pmatrix} \zeta_{1j} \\ \zeta_{2j} \end{pmatrix}$$

We used the `lme4` and `OpenMx` packages in the R environment (R Core Team 2013) to estimate the models (for the code: see Appendix). Results are shown in Table 8.1 (all estimates are significant at the $\alpha = 1\%$ level, unless otherwise indicated). We can see that SRH decreases slightly over time (β_{10} in LMEM notation, cf. Eq. 8.6; α_1 in LCM notation, cf. Eq. 8.7), and that women report slightly poorer health (β_{01} ; Γ_1). Furthermore, women and men have similar overall intraindividual change in SRH (β_{11} ; Γ_2 are non-significant). Satisfaction with personal relationships has a positive association with SRH (π_2 ; B), so that an increase of 1 point in satisfaction with relationships yields a 0.044 (0.043 in the LCM) increase in SRH. Finally, compared to the error variance (σ^2_E), there is a relatively large interindividual variability in SRH (σ^2_I), and a small variability in intraindividual change (σ^2_S). The covariance between the random intercept and random slope (correlation = -0.38) indicates that individuals with higher initial SRH have a tendency to decline faster (a larger negative slope) than individuals with lower SRH.

As can be seen from the results displayed in Table 8.1, the solutions estimated from the LMEM and LCM are virtually identical. The slight differences are not meaningful and can be imputed to differences in the optimization algorithms adopted by the two approaches.

Table 8.1 Estimates of LMEM and LCM to predict evolution of SRH over time with sex and satisfaction with relationships as independent variables

Independent variables	LMEM		LCM	
	Notation	Estimate	Notation	Estimate
Intercept	β_{00}	3.760	α_1	3.760
Sex (ref: male)	β_{01}	−0.135	Γ_1	−0.133
Time (for males)	β_{10}	−0.015	α_2	−0.015
Interaction time: sex	β_{11}	0.001 ^{ns}	Γ_2	0.0004 ^{ns}
Satisfaction with relationships	π_2	0.044	B	0.043
Intercept variance	σ^2_I	0.223	σ^2_I	0.227
Slope variance	σ^2_S	0.002	σ^2_S	0.002
Covariance intercept-slope	σ_{IS}	−0.009	σ_{IS}	−0.009
Error variance	σ^2_E	0.252	σ^2_E	0.253

Source of data: Swiss Household Panel
ns non-significant at $\alpha = 1\%$

Three Notable Extensions

Both the LMEM and the LCM have been extended to accommodate three notable features. First, because of the practical realities of any longitudinal study, a statistical model aimed at analyzing longitudinal data must be able to handle incomplete data. For instance, in any long-term longitudinal study lasting several weeks, months, years, or decades it is virtually impossible that all participants present at inception remain for the total duration. So-called micro-longitudinal studies (such as a psychological experiments lasting a few minutes, during which individuals are repeatedly assessed) are not necessarily immune to this type of dropout mechanism, either. This gives rise to data being incomplete for some, often most, participants.

Pretending data to be complete by limiting one’s analyses to the entities that provided all longitudinal assessments (so-called complete-case analyses) will most likely lead to biased estimation, because the subsample with complete data usually represents a selection of the starting sample. That is, the parameter estimates obtained from the application of any statistical model to complete-case longitudinal data will likely be different (biased) from the population parameters. Note that it is not the quantity of data incompleteness that is necessarily the chief determinant of the amount of estimation bias, but rather the nature of (the process causing) incomplete data.

Second, the frequent use of the LMEM and LCM to single variables assessed repeatedly has motivated multivariate applications to address relations in change processes. It is not realistic, for any substantive life course researcher, to theorize single behaviors or attitudes in isolation. Most often, from a theoretical perspective, it is most interesting to study interrelationships between multiple variables, to assess, for instance, if an entity changing on one behavior also tends to change on another behavior. Multivariate extensions of both the LMEM and the LCM easily address such questions.

Third, it is often illusory to assume that the samples at hand represent homogeneous groupings of entities. We may have reasons to believe that known characteristics of the entities may influence their change process. Males vs. females, wealthy vs. poor, with a low vs. a high level of education, etc. may differ from each other. Rather than considering such characteristics as covariates in the LMEM and LCM, it is possible to use them to create subsamples, which are then compared with respect to all parameters of the models. Moreover, group membership may not be known, but uncovered, in a rather exploratory fashion, from the data. This last application has received wide attention in several disciplines studying the life course.

Incomplete Data

Virtually every longitudinal study faces the reality of not having all assessments at all time points on all participants. For instance, in the illustration above, by the end of the study 50.4% of the initial participants were no longer in the study. Modern methods allow obtaining unbiased estimates despite the incompleteness of the outcome data Y . Under specific conditions (see Rubin 1976), any variable X that may be related to the reasons of data incompleteness in Y is called informative or auxiliary. For instance, older and less healthy participants are more likely to drop out of a longitudinal study than younger and healthier participants. Thus, age and health are informative X variables that are related to the probability of not providing outcome Y values. Informative variables can be added to a longitudinal model to decrease estimation bias. This is why it is of chief importance to measure as many potentially informative X variables as possible at the first wave of a longitudinal study, when attrition (i.e., incomplete data in Y) has not yet occurred. This information may be used in subsequent waves to lessen estimation bias (for more detail, see Graham 2009, and Schafer and Graham 2002). In the previous illustration, satisfaction with personal relationships was related to the probability of dropping out, so that its inclusion in the model does not only address an interesting substantive issue but also reduces overall estimation bias.

In some situations it is desirable to have incomplete data, mainly to avoid other, more deleterious effects. For instance, to understand the interrelationships among a very big set of variables, it is not necessary to administer all variables to all participants. Doing so could introduce major fatigue and demotivation effects, which are likely to lower the validity and reliability of the assessments. Rather, it is possible to randomly create subgroups of participants, each of which is administered a reduced set of variables. What matters is that the subsets of variables overlap across the participants, so that all interrelationships among the variables can be estimated. If carefully planned, then, the administration time and participants' burden can be reduced dramatically, without increase in parameter bias nor the need for informative variables (McArdle 1994).

In longitudinal research this feature may be of crucial importance. Researchers interested in wide segments of the life course (say, young adulthood, from age

18 to 35 years) can probably not afford to observe an age-homogeneous sample during the entire time epoch of interest (that is, study a sample of 18-year old participants during 17 years, until they are 35 years old). By the means of so-called age-convergence analyses (Bell 1953, 1954) and cohort sequential designs (Schaie 1977), it is possible to reduce the overall length of the study, without having to shorten the time epoch of substantive interest. For instance, three samples, of 18-, 23-, and 28-year old participants, may be observed for seven years. The total interval of time is now covered in seven rather than 17 years, and the partial age overlap of the three samples (18–25, 23–30, and 28–35 years) will allow testing whether the samples develop alike (i.e., convergence).

The software implementing the LMEM is more flexible than that of LCM analyses when dealing with incomplete Y data. Indeed, the optimization algorithms used to estimate parameters with incomplete data sets are usually much more efficient for the LMEM. Nevertheless, the LCM allows for a more flexible treatment of informative covariates, so that in problematic instances of incomplete data the LCM may be the preferred strategy (Graham 2003). However, neither the LMEM nor the LCM compensate directly for missing X values. In that case, two valuable alternative strategies are to either apply multiple imputation first, to impute all incomplete X and Y values, and then apply the longitudinal model (e.g., Carpenter and Goldstein 2004), or to estimate the model in the Bayesian framework (e.g., Muthén and Muthén 1998–2012).

Multivariate Extensions

By definition, a statistical model for longitudinal data must involve multiple variables, in that the outcome of interest must be observed at least twice (at times $i = 1$ and $i = 2$). Despite this technicality, it is common to call a longitudinal model applied to a single outcome *univariate*, denoting that a single Y outcome is analyzed. It follows that a *multivariate* longitudinal model is one that analyzes at least two outcomes (Y_1 and Y_2). In its simplest specification, a multivariate LMEM supposes an intercept and a slope growth factor for each outcome, which covary freely (MacCallum et al. 1997). This allows assessing the degree of communality of multiple change processes. Note that the functional form of growth need not be the same across the outcomes. In the previous illustration it would be possible to study change in both self-rated health and satisfaction with personal relationships. A multivariate LMEM might specify an intercept and a slope for both outcomes and estimate all six covariances among the four growth factors.

The LCM can also be easily adapted to multiple outcomes. Here, the association between the growth factors is not limited to symmetrical effects (i.e., covariances), but may be freely specified by the analyst. For instance, it is possible to let the growth factors of an earlier process predict those of a later process (rather than simply covary with them; Singer and Willett 2003). Such a relation cannot be tested within a multivariate LMEM. In the previous illustration we might want to assess

the effect of degree of satisfaction with personal relationship and change therein on self-rated health. Then, we might define a multivariate LCM with satisfaction assessments from 2004 to 2007 and health assessments from 2008 to 2011. We could then estimate the effects of the intercept and slope of the former (and temporally preceding) construct on the intercept and slope of the latter.

Another important multivariate extension of the LCM stems from its implementation within the structural equation modeling framework, where it is possible to define latent variables based on the common, shared variance of a chosen set of observed variables via a common factor model (Spearman 1904). Assuming the common factor can be defined at each wave of measurement of a longitudinal study, the outcome of interest is no longer an observed variable, but the latent variable itself. A LCM (with growth factors of higher order) can hence be specified to assess the growth of the common factor (McArdle called this extension a Curve of Factors Model; McArdle 1988). For instance, rather than relying on a single health question, multiple questions (and/or objective health measurements) could be assessed, to define a common health factor. The LCM would then study the change trajectory not of the single health assessments, but of the common health factor.

Specific multivariate extensions of the LMEM/LCM are particularly useful when trying to assess causality relations between multiple outcomes. For instance, the LCM can be modified to define multiple slope factors, each acting between two adjacent $i-1$ and i assessments. Then, each assessment at time $i-1$ can influence the immediately upcoming change between $i-1$ and i . In a bivariate setting it is then possible to estimate, for instance, the influence of variable A at time $i-1$ on the change in variable B between $i-1$ and i , and vice-versa. This analysis would allow determining whether satisfaction with personal relationships drives changes in self-assessed health, or vice-versa, or both variables influence each other's change. Similar extensions are possible within the multivariate LMEM, with the inclusion of instrumental variables to predict time-invariant and/or time-varying confounders (e.g., Skrondal and Rabe-Hesketh 2004). However, for this extension, as well as for all models discussed here, it is important to bear in mind that they cannot be considered as proofs of causal relationships, because it is not conceivable to assure that no other construct or variable is causing such relationships.

Multiple Group Analyses

The specifications of growth models discussed thus far assume that the sample is homogeneous with respect to the growth process, meaning that all entities within the sample follow the same functional form of growth (albeit allowing for interindividual differences with respect to the magnitude of the growth factors – i.e., random effects). Adding a time-invariant covariate that clearly splits the overall sample in subsamples (e.g., men and women), as discussed in Eqs. 8.4 and 8.5, allows for a difference in mean intercept and mean slope between men and women. However, all other parameters and the shape of change remain unaltered across the sexes. The adequacy of this constraint can be tested. A multiple-group analysis, by

sex, of the previous illustration would reveal that men and women have different mean intercepts of health (with women's inferior to men's), but similar mean slopes (consistent with the non-significant time-by-sex interaction in Table 8.1).

Splitting the sample into two subsamples allows freeing the equality constraints of all remaining parameters, as well as recovering group-specific growth features that would otherwise go unobserved. For instance, in a clinical trial we would expect a treatment group to react differently from a control group, not just with respect to the average intercept and slope mean, but much more generally. The control group may not undergo any change and remain constant. Hence, the growth model may be reduced to an intercept-only model. The treatment group would probably undergo marked change, as a result of the treatment, which would require the growth model to specify factors to this effect. Such a group difference cannot be modeled by considering group membership as a simple covariate.

In yet other cases we may suspect the existence of different subgroups, with respect to the analyzed change process, but ignore both the groups' characteristics and individual group membership. Thus, in a standard LCM, we cannot split the sample into subsamples according to the values of a known grouping variable. In this case, the grouping variable is assumed latent or unobserved, and we ignore its values. Recent developments extended the LCM to allow uncovering subgroups based on latent grouping variables. Latent class growth analyses (Nagin 1999) and growth mixture models (Muthén et al. 2002) are two such models, and both are currently only implemented in the LCM framework. Advanced versions of this approach can also be applied to state-trait type models, in order to uncover groups of individuals based on their degree of variability vs. stability (Courvoisier et al. 2007).

Conclusions

The LMEM and the LCM are statistical models of change that have permeated in many scientific disciplines. Today, they are largely used, thanks also to excellent reviews and textbooks, often supplemented by online material (e.g., Collins 2006; Duncan et al. 2006; McArdle and Nesselroade 2014; Pinheiro and Bates 2000; Singer and Willett 2003). Additionally, university curricula in psychology, sociology, demography, gerontology, and other disciplines related to the study of the life course have started integrating courses or workshops that discuss, among other statistical models, the LMEM and the LCM.

We believe that disciplines concerned with the study of the life course and the lifespan can greatly benefit from using the LMEM and the LCM. At the same time, we do not conceive statistics to be an independent academic discipline that mainly provides tools to substantive researchers. The many theoretical and methodological challenges of life course research call for flexible statistical tools. Ideally, statisticians and substantive life course researchers engage in dialogues and make progress together. Increased use of the LMEM and the LCM by life course researchers will motivate statisticians to additionally extend both models. With this chapter we hope to have further encouraged this dialogue.

Appendix

```
#####
# Ghisletta, P., Renaud, O., Jacot, N., and Courvoisier, D. (2014)
# Linear mixed-effects and latent curve models for
# longitudinal life course analyses.
# In Burton-Jeangros, C., Cullati, S., and Sacker, A. (Eds).
# A life course perspective on health trajectories and transitions
# Berlin, Germany: Springer
#
#
#       R code for analyses on Swiss Household Panel, 2004-2011
#
#####

# Import the data
shp.data <- read.table(file="shp04_l1.csv",
                      header=TRUE, sep=";")

# Select the variables of interest
shp <- shp.data[,c("IDPERS", "SEX",
                  "P04C01", "P05C01",
                  "P06C01", "P07C01",
                  "P08C01", "P09C01",
                  "P10C01", "P11C01",
                  "P04QL04", "P05QL04",
                  "P06QL04", "P07QL04",
                  "P08QL04", "P09QL04",
                  "P10QL04", "P11QL04")]

#####
# LMEM

#####
# Reshape data in a long format
shp.lmem <- reshape(shp, idvar="IDPERS", times=0:7,
                   direction="long",
                   varying=list(c("P04C01", "P05C01",
                                "P06C01", "P07C01",
                                "P08C01", "P09C01",
                                "P10C01", "P11C01"),
                              c("P04QL04", "P05QL04",
                                "P06QL04", "P07QL04",
                                "P08QL04", "P09QL04",
                                "P10QL04", "P11QL04")),
                   v.names=c("SRH", "SATREL"))

# Replace the values inferior to 1 by NA for the variable SRH
shp.lmem$SRH[shp.lmem$SRH<1] <- NA

# Replace the values inferior to 0 by NA for the variable SATREL
shp.lmem$SATREL[shp.lmem$SATREL<0] <- NA

# Load the package car that contains the function recode()
```

```

library(car)

# Recode the variable SRH in order to have
# 1=poor health and 5=excellent health
shp.lmem$SRH <- recode(shp.lmem$SRH, "1=5 ; 2=4; 3=3; 4=2; 5=1")

# Recode the variable sex in order to have 0=male and 1=female
shp.lmem$SEX <- shp.lmem$SEX-1

#####
# Estimation

# Load the package lme4 that contains the function lmer()
library(lme4)

# Fit the model
LMEM.model <- lmer(SRH ~ 1+ SEX*time + SATREL + (1+time|IDPERS),
                    data=shp.lmem, REML=FALSE)

# Summary statistics of the model
summary(LMEM.model)

#####
# LCM

#####
# Reshape data in a wide format
shp.lcm.data <- reshape(shp.lmem, idvar="IDPERS", timevar="time",
                        direction="wide",
                        v.names=c("SRH", "SATREL"))

# Load the package reshape that contains the function rename()
library(reshape)

# As the '.' is illegal for manifest variables in the function
# mxRun(), rename the variables SRH AND SATREL
shp.lcm <- rename(shp.lcm.data, c("SRH.0"="SRH0", "SRH.1"="SRH1",
                                "SRH.2"="SRH2", "SRH.3"="SRH3",
                                "SRH.4"="SRH4", "SRH.5"="SRH5",
                                "SRH.6"="SRH6", "SRH.7"="SRH7",
                                "SATREL.0"="SATREL0",
                                "SATREL.1"="SATREL1",
                                "SATREL.2"="SATREL2",
                                "SATREL.3"="SATREL3",
                                "SATREL.4"="SATREL4",
                                "SATREL.5"="SATREL5",
                                "SATREL.6"="SATREL6",
                                "SATREL.7"="SATREL7"))

#####
# Estimation

# Load the package OpenMx that contains the functions
# mxModel() and mxRun()
library(OpenMx)

# Specify the manifest variables' names

```

```

indic <- c("SRH0", "SRH1", "SRH2", "SRH3",
          "SRH4", "SRH5", "SRH6", "SRH7", "SEX",
          "SATREL0", "SATREL1", "SATREL2", "SATREL3",
          "SATREL4", "SATREL5", "SATREL6", "SATREL7")

# Create the model
LCM.model <- mxModel("LCM", type="RAM",
  manifestVars=indic,
  latentVars=c("B0", "B1"),
  mxPath(from="B0", to=c("SRH0", "SRH1",
                        "SRH2", "SRH3",
                        "SRH4", "SRH5",
                        "SRH6", "SRH7"),
    arrows=1, free=FALSE, values=1),
  mxPath(from="B1", to=c("SRH0", "SRH1",
                        "SRH2", "SRH3",
                        "SRH4", "SRH5",
                        "SRH6", "SRH7"),
    arrows=1, free=FALSE,
    values=c(0,1,2,3,4,5,6,7)),
  mxPath(from=c("SRH0", "SRH1",
                "SRH2", "SRH3",
                "SRH4", "SRH5",
                "SRH6", "SRH7"),
    arrows=2, free=TRUE, values=1,
    labels="Ve"),
  mxPath(from="one", to=c("B0", "B1", "SEX",
                        "SATREL0",
                        "SATREL1",
                        "SATREL2",
                        "SATREL3",
                        "SATREL4",
                        "SATREL5",
                        "SATREL6",
                        "SATREL7"),
    arrows=1, free=TRUE,
    labels=c("MB0", "MB1", "MSEX",
            "MSATREL0", "MSATREL1",
            "MSATREL2", "MSATREL3",
            "MSATREL4", "MSATREL5",
            "MSATREL6", "MSATREL7")),
  mxPath(from=c("B0", "B1", "SEX",
                "SATREL0", "SATREL1",
                "SATREL2", "SATREL3",
                "SATREL4", "SATREL5",
                "SATREL6", "SATREL7"),
    arrows=2, free=TRUE,
    values=c(1,1,1,1,1,1,1,1,1,1,1),
    labels=c("VB0", "VB1", "VSEX",
            "VSATREL0", "VSATREL1",
            "VSATREL2", "VSATREL3",
            "VSATREL4", "VSATREL5",
            "VSATREL6", "VSATREL7")),
  mxPath(from="B0", to="B1", arrows=2,
    free=TRUE, values=-0.0086,
    labels="CB0B1"),
  mxPath(from="SEX", to=c("B0", "B1"),
    arrows=1, free=TRUE,

```

```

        values=c(-0.1349,0.0010),
        labels=c("SEXB0","SEXB1")),
mxPath(from="SATREL0", to="SRH0",
        arrows=1, free=TRUE,
        values=0.0436,
        labels="SATRELB"),
mxPath(from="SATREL1", to="SRH1",
        arrows=1, free=TRUE,
        values=0.0436,
        labels="SATRELB"),
mxPath(from="SATREL2", to="SRH2",
        arrows=1, free=TRUE,
        values=0.0436,
        labels="SATRELB"),
mxPath(from="SATREL3", to="SRH3",
        arrows=1, free=TRUE,
        values=0.0436,
        labels="SATRELB"),
mxPath(from="SATREL4", to="SRH4",
        arrows=1, free=TRUE,
        values=0.0436,
        labels="SATRELB"),
mxPath(from="SATREL5", to="SRH5",
        arrows=1, free=TRUE,
        values=0.0436,
        labels="SATRELB"),
mxPath(from="SATREL6", to="SRH6",
        arrows=1, free=TRUE,
        values=0.0436,
        labels="SATRELB"),
mxPath(from="SATREL7", to="SRH7",
        arrows=1, free=TRUE,
        values=0.0436,
        labels="SATRELB"),
mxData(observed=shp.lcm, type="raw")
)

# Fit the model
LCM.fit <- mxRun(LCM.model)
# Warning message:
# In model 'LCM' NPSOL returned a non-zero status code 6.
# The model does not satisfy the first-order
# optimality conditions to the required accuracy, and no
# improved point for the merit function could
# be found during the final linesearch (Mx status RED)

# To take into account the warning message,
# fit again the model with starting values that are equal
# to the values estimated in LCM.fit
LCM.model.new <- LCM.fit
LCM.fit.new <- mxRun(LCM.model.new)

# Summary statistics of the model
summary(LCM.fit.new)

```

Open Access This chapter is distributed under the terms of the Creative Commons Attribution Noncommercial License, which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Baltes, P. B. (1987). Theoretical propositions of life-span developmental psychology: On the dynamics between growth and decline. *Developmental Psychology*, 23(5), 611–626.
- Baltes, P. B., & Nesselroade, J. R. (1979). History and rationale of longitudinal research. In J. R. Nesselroade & P. B. Baltes (Eds.), *Longitudinal research in the study of behavior and development* (pp. 1–39). New York: Academic.
- Bauer, D. J. (2003). Estimating multilevel linear models as structural equation models. *Journal of Educational and Behavioral Statistics*, 28, 135–167.
- Bell, R. Q. (1953). Convergence: An accelerated longitudinal approach. *Child Development*, 24, 145–152.
- Bell, R. Q. (1954). An experimental test of the accelerated longitudinal approach. *Child Development*, 25, 281–286.
- Berkman, L. F., & Syme, S. L. (1979). Social networks, host resistance, and mortality: A nine-year follow-up study of Alameda county residents. *American Journal of Epidemiology*, 109, 186–204.
- Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation approach*. Hoboken: Wiley.
- Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin*, 101, 147–158.
- Carpenter, J., & Goldstein, H. (2004). Multiple imputation in MLwiN. *Multilevel Modelling Newsletter*, 16(2), 9–18.
- Chou, C.-P., Bentler, P. M., & Pentz, M. A. (1998). Comparisons of two statistical approaches to study growth curves: The multilevel model and latent curve analysis. *Structural Equation Modeling*, 5, 247–266.
- Collins, L. M. (2006). Analysis of longitudinal data: the integration of theoretical model, temporal design, and statistical model. *Annual review of psychology*, 57, 505–528.
- Courvoisier, D. S., Eid, M., & Nussbeck, F. W. (2007). Mixture distribution latent state-trait analysis: Basic ideas and applications. *Psychological Methods*, 12, 80–104.
- Cullati, S., Rousseaux, E., Gabadinho, A., Courvoisier, D. S., & Burton-Jeangros, C. (2014). Factors of change and cumulative factors in self-rated health trajectories: A systematic review. *Advances in Life Course Research*, 19, 14–27. <http://doi.org/10.1016/j.alcr.2013.11.002>.
- Davidian, M., & Giltinan, D. M. (1995). *Nonlinear models for repeated measurement data*. London: Chapman & Hall.
- Duncan, T. E., Duncan, S. C., & Strycker, L. A. (2006). *An introduction to latent variable growth curve modeling* (2nd ed.). New York: Routledge Academic.
- Goldstein, H. (1989). Models for multilevel response variables with an application to growth curves. In R. D. Bock (Ed.), *Multilevel analysis of educational data* (pp. 107–125). San Diego: Academic.
- Goldstein, H. (2011). *Multilevel statistical models* (Vol. 4). West Sussex: John Wiley and Sons.
- Ghisletta, P., & Lindenberger, U. (2004). Static and dynamic longitudinal structural analyses of cognitive changes in old age. *Gerontology*, 50, 12–16.
- Graham, J. W. (2003). Adding missing-data-relevant variables to FIML-based structural equation models. *Structural Equation Modeling*, 10, 80–100.
- Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, 60, 549–576.

- Halford, C., Wallman, T., Welin, L., Rosengren, A., Bardel, A., Johansson, S., & Svärdsudd, K. (2012). Effects of self-rated health on sick leave, disability pension, hospital admissions and mortality. A population-based longitudinal study of nearly 15,000 observations among Swedish women and men. *BMC Public Health*, 12, 1103.
- Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, 38, 963–974.
- MacCallum, R. C., Kim, C., Malarkey, W. B., & Kiecolt-Glaser, J. K. (1997). Studying multivariate change using multilevel models and latent curve models. *Multivariate Behavioral Research*, 32, 215–253.
- McArdle, J. J. (1986). Latent growth within behavior genetic models. *Behavior Genetics*, 16, 163–200.
- McArdle, J. J. (1988). Dynamic but structural equation modeling of repeated measures data. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (pp. 561–614). New York: Plenum Press.
- McArdle, J. J. (1994). Structural factor analysis experiments with incomplete data. *Multivariate Behavioral Research*, 29, 409–454.
- McArdle, J. J. (2001). Growth curve analysis. In N. J. Smelser & P. B. Baltes (Eds.), *The international encyclopedia of the behavioral and social sciences* (pp. 6439–6445). New York: Pergamon Press.
- McArdle, J. J., & Hamagami, F. (1996). Multilevel models from a multiple group structural equation perspective. In G. A. Marcoulides & R. E. Schumaker (Eds.), *Advanced structural equation modeling. Issues and techniques* (pp. 89–124). Mahwah: Lawrence Erlbaum Associates.
- McArdle, J. J., & Hamagami, F. (2001). Latent difference score structural models for linear dynamic analyses with incomplete longitudinal data. In L. M. Collins & M. Sayer (Eds.), *New methods for the analysis of change* (pp. 139–175). Washington: American Psychological Association.
- McArdle, J. J., & Nesselroade, J. R. (2014). *Longitudinal data analysis using structural equation models*. Washington: American Psychological Association.
- McCulloch, C. E., & Searle, S. R. (2001). *Generalized, linear, and mixed models*. New York: Wiley.
- Mehta, P. D., & West, S. G. (2000). Putting the individual back into individual growth curves. *Psychological Methods*, 5, 23–43.
- Meredith, W., & Tisak, J. (1984). “Tuckerizing” curves. In *Annual meetings of psychometric society*. Santa Barbara.
- Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, 55, 107–122.
- Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus users's guide* (7th ed.). Los Angeles: Muthén and Muthén.
- Muthén, B. O., Brown, C. H., Masyn, K., Jo, B., Khoo, S.-T., Yang, C.-C., & Liao, J. (2002). General growth mixture modeling for randomized preventive interventions. *Biostatistics*, 3, 459–475.
- Nagin, D. S. (1999). Analyzing developmental trajectories: A semiparametric, group-based approach. *Psychological Methods*, 4, 139–157.
- Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effect models in S and S-PLUS*. New York: Springer.
- R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models. Applications and data analysis methods*. Thousand Oaks: Sage.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63, 581–592.
- Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7, 147–177.
- Schaie, K. W. (1977). Quasi-experimental research designs in the psychology of aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (pp. 39–57). New York: Van Nostrand Reinhold Company.
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis*. Oxford: Oxford University Press.

- Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*. Boca Raton: Chapman & Hall/CRC.
- Snijders, T. A. B., & Bosker, R. J. (1994). Modeled variance in two-level models. *Sociological Methods and Research*, 22, 342–363.
- Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). London: Sage.
- Spearman, C. (1904). General intelligence: Objectively determined and measured. *American Journal of Psychology*, 15, 201–293.
- Wainer, H. (2000). The centercept: An estimable and meaningful regression parameter. *Psychological Science*, 11, 434–436.